Journal of Multivariate Analysis 100 (2009) 377–386

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Testing for equality between two copulas

Bruno Rémillard [a,*], Olivier Scaillet [b,c]

[a] Service de l'enseignement des méthodes quantitatives de gestion, HEC Montréal, Montréal (Québec), Canada H3T 2A7

[b] HEC Université de Genève, Suisse

[c] Swiss Finance Institute, Faculté des SES, Bd Carl Vogt 102, CH - 1211 Geneve 4, Suisse

ARTICLE INFO

Article history:
Received 3 September 2007
Available online 15 May 2008

AMS 2000 subject classifications:
primary 60F05
secondary 62E20

JEL classification:
C12
C14
C15

Keywords:
Copula
Cramér–von Mises statistic
Empirical process
Pseudo-observations
Multiplier central limit theorem
p-value

ABSTRACT

We develop a test of equality between two dependence structures estimated through empirical copulas. We provide inference for independent or paired samples. The multiplier central limit theorem is used for calculating *p*-values of the Cramér–von Mises test statistic. Finite sample properties are assessed with Monte Carlo experiments. We apply the testing procedure on empirical examples in finance, psychology, insurance and medicine.

1. Introduction

Copulas are omnipresent in statistics and other fields like actuarial science, finance, reliability and hydrology, to name a few. This presence is explained by the copula being a summary of the full dependence structure between random variables. From a methodological point of view, most papers concentrate on parameter estimation, using ranks as in [14,24], or using estimated parametric margins, as in [19].

However functional nonparametric estimation of the copula has also been examined. It was first studied by Deheuvels in a series of papers [4–6,8,7], for the independent copula, and studied in full generality in [13]. Recent work on copula processes include [11,17]. Copula processes help in the development of tests for goodness-of-fit in semiparametric models, e.g. [10,15,16,23].

Another statistical issue related to copula modelling is the problem of testing for equality between two copulas. This yet unsolved issue aims at checking the validity of the hypothesis of two dependence structures being identical. For example, we could argue in credit risk that the copula of the joint default times of firms is the same as the copula of their respective asset values. See [9] for an illustration.

Our method to gauge the similarity between dependence structures has several advantages. First, it is applicable to any dimension. It is not restricted to the two dimensional case only. Second, it is not affected by strict monotonic transformations

* Corresponding author.
E-mail addresses: bruno.remillard@hec.ca (B. Rémillard), olivier.scaillet@hec.unige.ch (O. Scaillet).

0047-259X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.05.004

of the variables like log or exp transforms. Copulas possess an invariance property with respect to such mappings. This is a clear benefit over using a standard correlation to measure dependence. Third, it is model free. We rely on empirical estimation of copulas following a nonparametric approach. Fourth, finite sample properties are expected to be well behaved since we rely on a simulation strategy induced by a multiplier method. Our Monte Carlo results confirm this conjecture. The testing procedure performs well in samples as small as $n_1 = n_2 = 50$ and $d = 2$. Fifth, the test statistic takes a closed form. This improves the numerical speed of the simulation based testing procedure.

In this paper we illustrate the testing procedure on several empirical examples. We investigate questions arising in finance, psychology, insurance and medicine. The first application concerns the dependence structure between expense ratio and turnover level within two categories of US mutual funds. The second application examines the links between emotional experience and life satisfaction in the Chinese culture vis-à-vis the American culture. The third one is dedicated to the analysis of losses and allocated loss adjustment expenses (ALAEs, in short). In the last application, we investigate the dependence structure over time between two methods of assessment of depression. Other potential applications include investigating dependence between product sales in different retail stores (marketing), between income and consumption in different countries (economics), between reported items on corporate balance sheets in different countries (accounting), etc.

To describe the problem at hand, suppose we face two independent samples of $\mathbb{R}^d$-valued vectors. The first sample, $X_1, \ldots, X_{n_1}$ is taken from a distribution function F with continuous margins $F_1, \ldots, F_d$, and the second sample $Y_1, \ldots, Y_{n_2}$ is taken from a distribution function G with continuous margins $G_1, \ldots, G_d$. The vectors X_i, $i = 1, \ldots, n_1$, and Y_i, $i = 1, \ldots, n_2$, have size d, and entries denoted by X_{il} and Y_{il}, $l = 1, \ldots, d$. Then the unique copulas C and D associated with the first and second samples are determined, for any $x = (x_1, \ldots, x_d)$, by

$$F(x) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \qquad G(x) = D\{G_1(x_1), \ldots, G_d(x_d)\}.$$

The aim of the paper is to show how we can test the hypotheses

$$H_0 : C = D \quad \text{vs} \quad H_1 : C \neq D.$$

Obviously this is not equivalent to testing for $F = G$. Here we focus on the equality between the dependence structure as posited by $C = D$, leaving the behavior of the margins out of our field of interest. By construction our method is invariant with respect to strict monotonic transformations of the data.

To obtain consistent tests, we rely on a statistic based on the integrated square difference between the empirical copulas C_{n_1} and D_{n_2} defined for any $u = (u_1, \ldots, u_d) \in [0, 1]^d$ by

$$C_{n_1}(u) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{I}(U_{i,n_1} \le u) = \frac{1}{n_1} \sum_{i=1}^{n_1} \prod_{l=1}^{d} \mathbb{I}(U_{il,n_1} \le u_l),$$

and

$$D_{n_2}(u) = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbb{I}(V_{i,n_2} \le u) = \frac{1}{n_2} \sum_{i=1}^{n_2} \prod_{l=1}^{d} \mathbb{I}(V_{il,n_2} \le u_l),$$

where $U_{i,n_1} = (U_{i1,n_1}, \ldots, U_{id,n_1})$, $V_{i,n_2} = (V_{i1,n_2}, \ldots, U_{id,n_2})$, and for any $l \in \{1, \ldots, d\}$,

$$U_{il,n_1} = \frac{n_1}{n_1 + 1} F_{l,n_1}(X_{il}) = \operatorname{rank}(X_{il})/(n_1 + 1), \quad 1 \le i \le n_1,$$
$$V_{il,n_2} = \frac{n_2}{n_2 + 1} G_{l,n_2}(Y_{il}) = \operatorname{rank}(Y_{il})/(n_2 + 1), \quad 1 \le i \le n_2,$$

with

$$F_{l,n_1}(x_l) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{I}(X_{il} \le x_l) \quad \text{and} \quad G_{l,n_2}(x_l) = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbb{I}(Y_{il} \le x_l),$$

being the empirical distribution functions of $(X_{il})_{i=1}^{n_1}$ and $(Y_{il})_{i=1}^{n_2}$, respectively, defined for any $x_l \in \mathbb{R}$.

Test statistics for the equality between two copulas rely on functionals of the empirical process

$$\mathbb{E}_{n_1,n_2} = (C_{n_1} - D_{n_2}) \Big/ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

The asymptotic behavior of $\mathbb{E}_{n_1,n_2}$ is given in Section 2, together with a simulation based method for computing p-values. Some numerical results are given in Section 3 to illustrate the finite sample properties of the testing procedure. Section 4 is dedicated to empirical applications. The proof of the theoretical results are relegated to Appendix A while explicit expressions for calculating the simulated Cramér–von Mises test statistics are available in Appendix B of [21].

2. Test statistic and main results

If the mappings $u \mapsto \partial_{u_l} C(u)$ are continuous on $[0,1]^d$, then it is known, see, e.g., [13,27], that $\mathbb{C}_{n_1} = \sqrt{n_1}(C_{n_1} - C)$ converges weakly in $\mathcal{D}([0,1]^d)$ to a continuous centered Gaussian process $\mathbb{C}$, denoted by $\mathbb{C}_{n_1} \rightsquigarrow \mathbb{C}$, where $\mathbb{C}$ has the representation

$$\mathbb{C}(u) = \alpha(u) - \sum_{l=1}^{d} \beta_l(u_l)\partial_{u_l} C(u), \tag{1}$$

with

$$\alpha_{n_1}(u) = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \{\mathbb{I}(U_i \le u) - C(u)\} \rightsquigarrow \alpha(u),$$

$\beta_l(u_l) = \alpha(1,\ldots,1,u_l,1,\ldots,1)$, $1 \le l \le d$, and $U_i = (F_1(X_{i1}),\ldots,F_d(X_{id}))$. Note that the extra term $\sum_{l=1}^{d} \beta_l(u_l)\partial_{u_l} C(u)$ comes from the marginal distributions $F_1,\ldots,F_d$ being unknown.

Similarly, $\mathbb{D}_{n_2} = \sqrt{n_2}(D_{n_2} - D) \rightsquigarrow \mathbb{D}$ in $\mathcal{D}([0,1]^d)$ where $\mathbb{D}$ is a continuous centered Gaussian process represented by

$$\mathbb{D}(u) = \gamma(u) - \sum_{l=1}^{d} \delta_l(u_l)\partial_{u_l} D(u), \tag{2}$$

with

$$\gamma_{n_2}(u) = \frac{1}{\sqrt{n_2}} \sum_{i=1}^{n_2} \{\mathbb{I}(V_i \le u) - D(u)\} \rightsquigarrow \gamma(u),$$

$\delta_l(u_l) = \gamma(1,\ldots,1,u_l,1,\ldots,1)$, $1 \le l \le d$, and $V_i = (G_1(Y_{i1}),\ldots,G_d(Y_{id}))$.

If $\min(n_1,n_2) \to \infty$, in such a way that $n_1/(n_1+n_2) \to \lambda \in [0,1]$, then (see the proofs of the theorems below)

$$\mathcal{E}_{n_1,n_2} = \sqrt{\frac{n_2}{n_1+n_2}}\mathbb{C}_{n_1} - \sqrt{\frac{n_1}{n_1+n_2}}\mathbb{D}_{n_2} \rightsquigarrow \mathcal{E} = \sqrt{1-\lambda}\,\mathbb{C} - \sqrt{\lambda}\,\mathbb{D}.$$

Under the null hypothesis $H_0: C = D$, we have $\mathbb{E}_{n_1,n_2} = \mathcal{E}_{n_1,n_2}$, and thus $\mathbb{E}_{n_1,n_2} \rightsquigarrow \mathcal{E}$.

To test the null hypothesis $H_0: C = D$, we propose to use the Cramér–von Mises principle, and build

$$\begin{aligned} S_{n_1,n_2} &= \int_{[0,1]^d} \mathbb{E}^2_{n_1,n_2}(u)\mathrm{d}u \\ &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \times \left\{ \frac{1}{n_1^2} \sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\prod_{s=1}^{d}\left(1 - U_{is,n_1} \vee U_{js,n_1}\right) \right. \\ &\quad \left. - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\prod_{s=1}^{d}\left(1 - U_{is,n_1} \vee V_{js,n_2}\right) + \frac{1}{n_2^2} \sum_{i=1}^{n_2}\sum_{j=1}^{n_2}\prod_{s=1}^{d}\left(1 - V_{is,n_2} \vee V_{js,n_2}\right) \right\}, \end{aligned}$$

where $a \vee b$ stands for $\max(a,b)$. When $C = D$, then

$$S_{n_1,n_2} \rightsquigarrow S = \int_{[0,1]^d} \mathcal{E}^2(u)\mathrm{d}u,$$

while if $C \ne D$, then $S_{n_1,n_2} \xrightarrow{Pr} \infty$. This yields consistency of the testing procedure.

Since C and D are unknown, computing p-values appears difficult at first sight. However, due to a powerful multiplier technique, we can estimate the p-value via simulations. In a single copula context the idea is already suggested in [22], and further developed in [20]. The trick is to use a multiplier central limit theorem [28] to approximate each random term appearing in (1) and (2). Note that a bootstrap approach would be inappropriate here since it fails to deliver consistency when applied to Cramér–von Mises test statistics (see Example 7 of [2], [1,3]).

To see how it works, suppose that for any $k \in \{1,\ldots,N\}$, $\xi_1^{(k)},\ldots,\xi_{n_1}^{(k)}, \zeta_1^{(k)},\ldots,\zeta_{n_2}^{(k)}$ are independent and identically distributed variables with mean zero and variance one.

Set

$$\begin{aligned} \hat{\alpha}^{(k)}_{n_1}(u) &= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \xi_i^{(k)} \left\{\mathbb{I}(U_{i,n_1} \le u) - C_{n_1}(u)\right\} \\ &= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \left(\xi_i^{(k)} - \bar{\xi}^{(k)}\right) \mathbb{I}(U_{i,n_1} \le u), \end{aligned}$$

$$\hat{\gamma}^{(k)}_{n_2}(u) = \frac{1}{\sqrt{n_2}} \sum_{i=1}^{n_2} \left(\zeta_i^{(k)} - \bar{\zeta}^{(k)}\right) \mathbb{I}(V_{i,n_2} \le u),$$

where $\bar{\xi}^{(k)} = \frac{1}{n_1}\sum_{i=1}^{n_1} \xi_i^{(k)}$, $\bar{\zeta}^{(k)} = \frac{1}{n_2}\sum_{i=1}^{n_2} \zeta_i^{(k)}$, and for any $l \in \{1, \ldots, d\}$,

$$\begin{aligned}\hat{\beta}_{l,n_1}^{(k)}(u_l) &= \hat{\alpha}_{n_1}^{(k)}(1, \ldots, 1, u_l, 1, \ldots, 1)\\ &= \frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\left(\xi_i^{(k)} - \bar{\xi}^{(k)}\right)\mathbb{I}(U_{il,n_1} \le u_k),\\ \hat{\delta}_{l,n_2}^{(k)}(u_l) &= \hat{\gamma}_{n_2}^{(k)}(1, \ldots, 1, u_l, 1, \ldots, 1)\\ &= \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}\left(\zeta_i^{(k)} - \bar{\zeta}^{(k)}\right)\mathbb{I}(V_{il,n_2} \le u_k).\end{aligned}$$

To approximate the partial derivatives ∇C and ∇D, we proceed as in [17]. For any $l \in \{1, \ldots, d\}$, set

$$\widehat{\partial_{u_l} C_{n_1,h_1}}(u) = \frac{C_{n_1}(u + h_1 e_l) - C_{n_1}(u - h_1 e_l)}{2h_1}$$

and

$$\widehat{\partial_{u_l} D_{n_2,h_2}}(u) = \frac{D_{n_2}(u + h_2 e_l) - D_{n_2}(u - h_2 e_l)}{2h_2},$$

where e_l is the lth column of the $d \times d$ identity matrix. We could also rely on a kernel based estimate of the derivative [12], but this would impede the writing of explicit expressions for the simulated test statistic, and slow down the procedure. These expressions are available on request from the authors, and also written down in Appendix B of [21].

Finally, for all $u \in [0, 1]^d$, and for all $k \in \{1 \ldots, N\}$, let

$$\hat{\mathbb{C}}_{n_1,h_1}^{(k)}(u) = \hat{\alpha}_{n_1}^{(k)}(u) - \sum_{l=1}^{d} \hat{\beta}_{l,n_1}^{(k)}(u_l)\widehat{\partial_{u_l} C_{n_1,h_1}}(u),$$

$$\hat{\mathbb{D}}_{n_2,h_2}^{(k)}(u) = \hat{\gamma}_{n_2}^{(k)}(u) - \sum_{l=1}^{d} \hat{\delta}_{l,n_2}^{(k)}(u_l)\widehat{\partial_{u_l} D_{n_2,h_2}}(u),$$

and

$$\hat{\mathcal{E}}_{n_1,n_2}^{(k)} = \sqrt{\frac{n_2}{n_1+n_2}}\hat{\mathbb{C}}_{n_1,h_1}^{(k)} - \sqrt{\frac{n_1}{n_1+n_2}}\hat{\mathbb{D}}_{n_2,h_2}^{(k)}.$$

Further set

$$S_{n_1,n_2}^{(0)} = \int_{[0,1]^d} \mathcal{E}_{n_1,n_2}^2(u)\mathrm{d}u$$

and

$$\hat{S}_{n_1,n_2}^{(k)} = \int_{[0,1]^d} \left\{\hat{\mathcal{E}}_{n_1,n_2}^{(k)}\right\}^2(u)\mathrm{d}u, \quad k \in \{1, \ldots, N\}.$$

Theorem 2.1 (*Independent Samples*). *Suppose that ∇C and ∇D are continuous on $[0, 1]^d$. If $h_i = n_i^{-1/2}$, $i = 1, 2$ and if $\min(n_1, n_2) \to \infty$ in such a way that $n_1/(n_1 + n_2) \to \lambda \in (0, 1)$, then*

$$\left(\mathcal{E}_{n_1,n_2}, \hat{\mathcal{E}}_{n_1,n_2}^{(1)}, \ldots, \hat{\mathcal{E}}_{n_1,n_2}^{(N)}\right) \rightsquigarrow \left(\mathcal{E}, \tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)}\right) \quad \text{in } \mathcal{D}\left([0, 1]^d\right)^{\otimes(N+1)},$$

where $\tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)}$ are independent copies of $\mathcal{E}$. In particular,

$$\left(S_{n_1,n_2}^{(0)}, \hat{S}_{n_1,n_2}^{(1)}, \ldots, \hat{S}_{n_1,n_2}^{(N)}\right) \rightsquigarrow \left(S, \tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}\right) \quad \text{in } [0, \infty)^{\otimes(N+1)},$$

where $\tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}$ are independent copies of $S = \int_{[0,1]^d} \mathcal{E}^2(u)\mathrm{d}u$. An approximate p-value for S_{n_1,n_2} is then given by

$$\frac{1}{N}\sum_{k=1}^{N}\mathbb{I}\left(\hat{S}_{n_1,n_2}^{(k)} > S_{n_1,n_2}\right).$$

The proof is given in Appendix A.1.

The previous theorem holds true for two independent populations. What about paired observations, i.e., X_i is not independent of Y_i, but $n_2 = n_1 = n$? It is easy to check that the previous methodology applies, provided we draw $\xi_i^{(k)}$ and set $\zeta_i^{(k)} = \xi_i^{(k)}$, for all $i = 1, \ldots, n$, and all $k = 1, \ldots, N$. In the next theorem we shorten the subscript ${}_{n,n}$ as ${}_n$.

Table 1
Size and power of the Cramér–von Mises test based on a multiplier technique with $N = 1000$, when $n_1 = 50, 100$, $n_2 = 50, 100$, $d = 2$, and Clayton copulas parameterized such that the Kendall tau is $\tau_C = 0.2$ for C, and $\tau_D = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ for D

(n_1, n_2)	Kendall tau τ_D	0.2	0.3	0.4	0.5	0.6	0.7	0.8
(50, 50)	Power (%)	4.9	9.4	28.9	58	87.4	97.6	99.9
(50, 100)	Power (%)	4.6	12.7	37.2	73.6	95.4	99.6	100
(100, 50)	Power (%)	5.4	14.3	40.3	74.4	95.7	99.9	100
(100, 100)	Power (%)	4.5	13.5	53.1	88.5	99.2	100	100

The significance level is 5%, and empirical levels are computed with 1000 replicates.

Theorem 2.2 (*Paired Samples*). *Suppose that ∇C and ∇D are continuous on $[0, 1]^d$. If $h_i = h = n^{-1/2}$, $i = 1, 2$ and if $n \to \infty$, then*

$$\left(\mathcal{E}_n, \hat{\mathcal{E}}_n^{(1)}, \ldots, \hat{\mathcal{E}}_n^{(N)}\right) \rightsquigarrow \left(\mathcal{E}, \tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)}\right) \quad \text{in } \mathcal{D}\left([0, 1]^d\right)^{\otimes(N+1)},$$

where $\tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)}$ are independent copies of $\mathcal{E}$. In particular,

$$\left(S_n^{(0)}, \hat{S}_n^{(1)}, \ldots, \hat{S}_n^{(N)}\right) \rightsquigarrow \left(S, \tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}\right) \quad \text{in } [0, \infty)^{\otimes(N+1)},$$

where $\tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}$ are independent copies of $S = \int_{[0,1]^d} \mathcal{E}^2(u)\mathrm{d}u$. An approximate p-value for S_n is then given by

$$\frac{1}{N} \sum_{k=1}^{N} \mathbb{I}\left(\hat{S}_n^{(k)} > S_n\right).$$

The proof is given in Appendix A.3.

3. Numerical experiments

From Theorem 2.1 we know that the level of the test should be correct when $n_1, n_2 \to \infty$. Here we check the finite sample properties of the testing procedure in terms of size and power. For the numerical experiments, the level of the test is fixed at 5%, so the power is estimated by the proportion of samples with p-value less than 5%. To this end, we have chosen three bivariate copula families (Clayton, Frank and Gumbel), all indexed by the Kendall tau $\tau(\theta)$ depending on the copula parameter θ. Recall that the Clayton copula is defined by all $u, v \in (0, 1)$ and parameter $\theta > 0$ by

$$C_\theta(u, v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}.$$

Here $\tau(\theta) = \theta/(\theta + 2)$.

The Frank copula is defined for all $u, v \in (0, 1)$ and $\theta > 0$ by

$$C_\theta(u, v) = \log\left(\frac{\theta + \theta^{u+v} - \theta^u - \theta^v}{\theta - 1}\right) \Big/ \log(\theta).$$

Then $\tau(\theta) = \frac{\log(\theta)^2 + 4\log(\theta) + 4\,\mathrm{dilog}(\theta)}{\log(\theta)^2}$, where $\mathrm{dilog}(x) = \int_1^x \frac{\log t}{1-t}\mathrm{d}t$.

Finally, the Gumbel copula is defined for all $u, v \in (0, 1)$ and $0 < \theta < 1$ by

$$C_\theta(u, v) = \exp\left[-\{(-\log u)^{1/\theta} + (-\log v)^{1/\theta}\}^\theta\right],$$

which gives $\tau(\theta) = 1 - \theta$.

As we can see from Table 1 for Clayton copulas, even for sample sizes as small as $n_1 = n_2 = 50$, the empirical level of the test (4.9%) is close to the theoretical one (5%). Moreover, the power of the test increases as expected, when D goes away from C, i.e., when τ_D increases, τ_C being fixed. It is close to 100% when τ_D is above.7 and τ_C is kept equal to.2. These results are confirmed by Table 2 for the Frank copula and by Table 3 for the Gumbel copula. Similar results also hold true for the other pairs of sample sizes $(n_1, n_2) = (50, 100), (100, 50), (100, 100)$.

4. Empirical applications

In this section we illustrate the testing procedures on empirical examples in finance, psychology, insurance and medicine. A generic Matlab code and its C add-in are available upon request from the authors for applied work. We have used $N = 1000$.

Table 2
Size and power of the Cramér–von Mises test based on a multiplier technique with $N = 1000$, when $n_1 = 50, 100, n_2 = 50, 100, d = 2$, and Frank copulas parameterized such that the Kendall tau is $\tau_C = 0.2$ for C, and $\tau_D = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ for D

(n_1, n_2)	Kendall tau τ_D	0.2	0.3	0.4	0.5	0.6	0.7	0.8
(50, 50)	Power (%)	4.7	10	32.9	55.9	89.9	99.1	99.9
(50, 100)	Power (%)	5.7	12.9	36.9	72.1	96.6	99.8	100
(100, 50)	Power (%)	4.8	15	45.1	74.9	98	100	100
(100, 100)	Power (%)	4.4	16.3	59.2	89.4	99.8	100	100

The significance level is 5%, and empirical levels are computed with 1000 replicates.

Table 3
Size and power of the Cramér–von Mises test based on a multiplier technique with $N = 1000$, when $n_1 = 50, 100, n_2 = 50, 100, d = 2$, and Gumbel copulas parameterized such that the Kendall tau is $\tau_C = 0.2$ for C, and $\tau_D = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ for D

(n_1, n_2)	Kendall tau τ_D	0.2	0.3	0.4	0.5	0.6	0.7	0.8
(50, 50)	Power (%)	4.2	8.4	26.5	57.1	85.6	98.5	99.9
(50, 100)	Power (%)	4.9	10.3	36.6	70.4	94.1	99.9	100
(100, 50)	Power (%)	4.6	14.7	39.4	73.1	96.4	99.9	100
(100, 100)	Power (%)	4.7	16.4	53.1	87.8	99.8	100	100

The significance level is 5%, and empirical levels are computed with 1000 replicates.

4.1. Expense ratio and turnover level

The data set is made of expense ratio and turnover level reported by 222 "Growth and Income" funds and 333 "Aggressive Growth" funds at the end of year 1994 (see, e.g., [30] for a detailed description of the data). A higher turnover induces higher transaction costs, and funds charge expenses partly to cover these costs. In 1994 growth-oriented funds maintain roughly 90% of their portfolios in equities, while income-oriented funds maintain a lower proportion around 80%. We want to study whether funds having different investment objectives share the same link between turnover level and expense ratio. The p-value is 0.425, and we conclude that the null hypothesis of equal dependence structure is not rejected at a 5% level. This means that the two categories of funds act in a similar way when adjusting the expenses they charge to recover their transaction costs.

4.2. Emotional experience and life satisfaction

The data set consists of positive affect scores (positive emotional mood) and life satisfaction scores (subjective well-being) recorded in China (559 university students) and the United States (443 university students) in the early 90s. We refer to the paper of [26] for data description and background on the psychological concepts. The question is whether the dependence structure for a collectivist culture, i.e., where a significant part of one's identity is made of collective elements, and that for an individualistic culture, i.e., where one's internal attributes are emphasized over the evaluations and expectations of others, can be considered to be equal or not. The p-value is 0, and we conclude that the null hypothesis of equal dependence structure is rejected at a 5% level. Hence the underlying culture has a significant impact.

4.3. Losses and ALAEs

Often actuaries have to price insurance contracts involving pairs of dependent variables. A classical example consists of computing the premium of a reinsurance treaty on a policy with unlimited liability, some retention level of the losses and a prorata sharing of ALAEs. Here ALAEs are types of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers' fees and claims' investigation expenses. The data are extracted from a database about medical insurance claims available from the Society of Actuaries. A thorough description of the data can be found in the monograph [18]. We analyze the dependence structure between losses (hospital charges) and ALAEs (other charges) for dependent females (967 observations) versus employee females (1116 observations) aged 30–39 in 1991 and insured by a Preferred Provider Organization (PPO) plan. The p-value is 0.065, and the null hypothesis of equal dependence structure is not rejected at a 5% level. We conclude that the status of the policy holder is irrelevant here (at a 5% level), and that premiums charged for both types of individuals should be the same if margins are roughly identical.

4.4. St John's wort versus sertaline

In [29] the authors compare the change in severity of depressive symptoms and occurrence of side effects in primary care patients treated with St John's wort and sertaline using a double-blind randomized 12-week trial. For each of the two treatment groups, depression was measured every two weeks with two different instruments: Hamilton raring scale for Depression (Ham-D) and Beck Depression Inventory (BDI). The authors conclude that there is no significant difference between the two treatments. By looking at the two groups, we now ask whether there is no change on the dependence structure of the two measures of depression over time. To this end, we use the methodology developed for paired samples. All ten pairs of measures corresponding to weeks 2, 4, 6, 8, 10 are compared, and we find that the largest estimated p-value is 0.001. Thus we have that the null hypothesis of equal dependence structure is rejected at a 5% level. This rejection might have an impact on the conclusion on the no difference between the two treatments since the relationship between the two measurement instruments is not the same in the two groups.

Acknowledgments

We would like to thank the associate editor and the three referees for constructive criticism and helpful comments. Partial funding in support of this work was provided by the Natural Sciences and Engineering Research Council of Canada, by the Fonds québécois de la recherche sur la nature et les technologies, as well as by the Institut de finance mathématique de Montréal and by the Swiss NSF through the NCCR Finrisk.

Appendix. Proofs of the results

Let $\xi_1, \ldots, \xi_n$ be independent and identically distributed random variables with mean zero and variance one. Also suppose that $X_1, \ldots, X_n$ are independent random vectors with continuous marginals $F_1, \ldots, F_d$ and copula C. Set $U_{ij} = F_j(X_{ij})$, $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, d\}$.

Then, for any $u = (u_1, \ldots, u_d) \in (0, 1)^d$, α_n and C_n can be expressed as

$$\alpha_n(u) = \sqrt{n}\left\{\tilde{C}_n(u) - C(u)\right\},$$

with

$$\tilde{C}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(U_{i1} \le u_1, \ldots, U_{id} \le u_d),$$

and

$$\begin{aligned} C_n(u) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(F_{n1}(X_{i1}) \le u_1, \ldots, F_{nd}(X_{id}) \le u_d\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(U_{i1} \le E_{n1}^{-1}(u_1), \ldots, U_{id} \le E_{nd}^{-1}(u_d)\right) \\ &= \tilde{C}_n\left(E_{n1}^{-1}(u_1), \ldots, E_{nd}^{-1}(u_d)\right), \end{aligned}$$

where for any $j \in \{1, \ldots, d\}$,

$$E_{nj}(u_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(U_{ij} \le u_j), \quad u_j \in [0, 1].$$

Furthermore, for any $u = (u_1, \ldots, u_d) \in [0, 1]^d$, set

$$\tilde{\alpha}_n(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i \left\{\mathbb{I}\left(U_{i1} \le u_1, \ldots, U_{id} \le u_d\right) - \tilde{C}_n(u)\right\}.$$

Then, for any $u = (u_1, \ldots, u_d) \in [0, 1]^d$,

$$\begin{aligned} \hat{\alpha}_n(u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i \left[\mathbb{I}\left(F_{n1}(X_{i1}) \le u_1, \ldots, F_{nd}(X_{id}) \le u_d\right) - C_n(u)\right] \\ &= \tilde{\alpha}_n\left(E_{n1}^{-1}(u_1), \ldots, E_{nd}^{-1}(u_d)\right). \end{aligned}$$

It follows from the classical multiplier central limit theorem [28] that $(\alpha_n, \tilde{\alpha}_n) \rightsquigarrow (\alpha, \tilde{\alpha})$ in $\mathcal{D}([0, 1]^d) \times \mathcal{D}([0, 1]^d)$, where $\tilde{\alpha}$ is an independent copy of α, and α is a C-Brownian bridge.

Next, since for any $j \in \{1, \ldots, d\}$, $\sup_{u_j \in [0,1]} \left|E_{nj}^{-1}(u_j) - u_j\right| = \sup_{u_j \in [0,1]} \left|E_{nj}(u_j) - u_j\right| \to 0$ as $n \to \infty$, e.g., [25], the following result holds.

Lemma A.1. $\left(\alpha_n, \hat{\alpha}_n\right) \rightsquigarrow (\alpha, \tilde{\alpha})$ *in* $\mathcal{D}([0, 1]^d) \times \mathcal{D}([0, 1]^d)$*, where* $\tilde{\alpha}$ *is an independent copy of* α*, and* α *is a* C*-Brownian bridge.*

A.1. Proof of Theorem 2.1

Proof. The proof is closely related to the one in [22]. Here we can simply use Lemma A.1 to conclude that, as $n_1 \to \infty$,

$$\left(\alpha_{n_1}, \hat{\alpha}_{n_1}^{(1)}, \ldots, \hat{\alpha}_{n_1}^{(N)}\right) \rightsquigarrow \left(\alpha, \tilde{\alpha}^{(1)}, \ldots, \tilde{\alpha}^{(N)}\right)$$

in $\mathcal{D}([0,1]^d)^{\otimes(N+1)}$, where $\tilde{\alpha}^{(1)}, \ldots, \tilde{\alpha}^{(N)}$ are independent copies of α, and α is a C-Brownian bridge.

Also, as $n_2 \to \infty$,

$$\left(\gamma_{n_2}, \hat{\gamma}_{n_2}^{(1)}, \ldots, \hat{\gamma}_{n_2}^{(N)}\right) \rightsquigarrow \left(\gamma, \tilde{\gamma}^{(1)}, \ldots, \tilde{\gamma}^{(N)}\right)$$

in $\mathcal{D}([0,1]^d)^{\otimes(N+1)}$, where $\tilde{\gamma}^{(1)}, \ldots, \tilde{\gamma}^{(N)}$ are independent copies of γ, and γ is a D-Brownian bridge.

As a consequence of independence between

$$\left(\alpha_{n_1}, \hat{\alpha}_{n_1}^{(1)}, \ldots, \hat{\alpha}_{n_1}^{(N)}\right) \quad \text{and} \quad \left(\gamma_{n_2}, \hat{\gamma}_{n_2}^{(1)}, \ldots, \hat{\gamma}_{n_2}^{(N)}\right),$$

we may conclude that as $\min(n_1, n_2) \to \infty$,

$$\left(\alpha_{n_1}, \gamma_{n_2}, \hat{\alpha}_{n_1}^{(1)}, \hat{\gamma}_{n_2}^{(1)}, \ldots, \hat{\alpha}_{n_1}^{(N)}, \hat{\gamma}_{n_2}^{(N)}\right) \rightsquigarrow \left(\alpha, \gamma, \tilde{\alpha}^{(1)}, \tilde{\gamma}^{(1)}, \ldots, \tilde{\alpha}^{(N)}, \tilde{\gamma}^{(N)}\right)$$

in $\mathcal{D}([0,1]^d)^{\otimes 2(N+1)}$, where $\left(\tilde{\alpha}^{(1)}, \tilde{\gamma}^{(1)}\right), \ldots, \left(\tilde{\alpha}^{(N)}, \tilde{\gamma}^{(N)}\right)$ are independent copies of (α, γ), α is independent of γ.

Next, since the conditions of Proposition A.2 of the next section are met, we obtain that for any $l \in \{1, \ldots, d\}$, $\widehat{\partial_{u_l} C_{n_1,h_1}}$ and $\widehat{\partial_{u_l} D_{n_2,h_2}}$ converge uniformly in probability to $\partial_{u_l} C$ and $\partial_{u_l} D$.

Hence $(\mathcal{E}_{n_1,n_2}, \hat{\mathcal{E}}_{n_1,n_2}^{(1)}, \ldots, \hat{\mathcal{E}}_{n_1,n_2}^{(N)}) \rightsquigarrow (\mathcal{E}, \tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)})$ in $\mathcal{D}\left([0,1]^d\right)^{\otimes(N+1)}$, where $\tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)}$ are independent copies of $\mathcal{E}$. Since the mapping $g \mapsto \int_{[0,1]^d} g^2(u)\mathrm{d}u$ is continuous, whenever g is continuous on $[0,1]^d$, it follows that

$$\left(S_{n_1,n_2}^{(0)}, \hat{S}_{n_1,n_2}^{(1)}, \ldots, \hat{S}_{n_1,n_2}^{(N)}\right) \rightsquigarrow \left(S, \tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}\right) \text{ in } [0,\infty)^{\otimes(N+1)},$$

where $\tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}$ are independent copies of $S = \int_{[0,1]^d} \mathcal{E}^2(u)\mathrm{d}u$. An approximate p-value for S_{n_1,n_2} is then given by $\frac{1}{N}\sum_{k=1}^N \mathbb{I}(\hat{S}_{n_1,n_2}^{(k)} > S_{n_1,n_2})$. $\square$

A.2. Uniform convergence of partial derivative estimates

Proposition A.2. *Suppose that* ∇C *and* ∇D *are continuous on* $[0,1]^d$. *Take* $h_i = n_i^{-1/2}$, $i = 1, 2$. *Then, as* $\min(n_1, n_2) \to \infty$,

$$\max_{1\le l\le d} \sup_{u\in[0,1]^d} \left|\widehat{\partial_{u_l} C_{n_1,h_1}}(u) - \partial_{u_l} C(u)\right| \stackrel{Pr}{\to} 0$$

and

$$\max_{1\le l\le d} \sup_{u\in[0,1]^d} \left|\widehat{\partial_{u_l} D_{n_2,h_2}}(u) - \partial_{u_l} D(u)\right| \stackrel{Pr}{\to} 0.$$

Proof. Let $l \in \{1, \ldots, d\}$ be fixed. Then,

$$\begin{aligned}\widehat{\partial_{u_l} C_{n_1,h_1}}(u) &= \frac{C_{n_1}(u + h_1 e_l) - C_{n_1}(u - h_1 e_l)}{2h_1} \\ &= \frac{C(u + h_1 e_l) - C(u - h_1 e_l)}{2h_1} + \frac{\mathbb{C}_{n_1}(u + h_1 e_l) - \mathbb{C}_{n_1}(u - h_1 e_l)}{2h_1\sqrt{n_1}}.\end{aligned}$$

Therefore we get by choosing $h_1 = n_1^{-1/2}$:

$$\begin{aligned}\sup_{u\in[0,1]^d} \left|\widehat{\partial_{u_l} C_{n_1,h_1}}(u) - \partial_{u_l} C(u)\right| &= \sup_{u\in[0,1]^d} \left|\frac{C_{n_1}(u + h_1 e_l) - C_{n_1}(u - h_1 e_l)}{2h_1} - \partial_{u_l} C(u)\right| \\ &\le \sup_{u\in[0,1]^d} \left|\frac{C(u + h_1 e_l) - C(u - h_1 e_l)}{2h_1} - \partial_{u_l} C(u)\right| \\ &\quad + \frac{1}{2} \sup_{u\in[0,1]^d} \left|\mathbb{C}_{n_1}(u + h_1 e_l) - \mathbb{C}_{n_1}(u - h_1 e_l)\right|,\end{aligned}$$

which tends to 0 as $n_1 \to \infty$, since $\partial_{u_l} C(u)$ is assumed to be continuous on $[0,1]^d$, and $\mathbb{C}_{n_1}$ converges in law to a continuous centered Gaussian process $\mathbb{C}$. The proof for $\widehat{\partial_{u_l} D_{n_2,h_2}}$ is similar. $\square$

A.3. Proof of Theorem 2.2

Proof. The proof is similar to the proof of Theorem 2.1. First, consider the independent vectors $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$, having copula $\mathcal{C}$ on $[0,1]^{2d}$, with the property that for any $u, v \in [0,1]^d$, $\mathcal{C}(u, 1, \ldots, 1) = C(u)$ and $\mathcal{C}(1, \ldots, 1, v) = D(v)$.

Next, for all $u, v \in [0,1]^d$, define

$$\mathcal{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(U_{i,n} \le u, V_{i,n} \le v),$$

$$\upsilon_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(U_i \le u, V_i \le v),$$

and

$$\hat{\upsilon}_n^{(k)}(u,v) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i^{(k)} \left\{\mathbb{I}(U_{i,n} \le u, V_{i,n} \le v) - \mathcal{C}_n(u,v)\right\}.$$

It follows from Lemma A.1 that, as $n \to \infty$,

$$\left(\upsilon_n, \hat{\upsilon}_n^{(1)}, \ldots, \hat{\upsilon}_n^{(N)}\right) \rightsquigarrow \left(\upsilon, \tilde{\upsilon}^{(1)}, \ldots, \tilde{\upsilon}^{(N)}\right)$$

in $\mathcal{D}([0,1]^d)^{\otimes(N+1)}$, where $\tilde{\upsilon}^{(1)}, \ldots, \tilde{\upsilon}^{(N)}$ are independent copies of υ, and υ is a $\mathcal{C}$-Brownian bridge.

Since for any $u, v \in [0,1]$, we have

$$C_n(u) = \mathcal{C}(u, 1, \ldots, 1), \qquad \tilde{\alpha}_n(u) = \upsilon_n(u, 1, \ldots, 1), \qquad \hat{\alpha}_n^{(k)}(u) = \hat{\upsilon}_n^{(k)}(u, 1, \ldots, 1)$$

and

$$D_n(u) = \mathcal{C}(1, \ldots, 1, v), \qquad \tilde{\gamma}_n(v) = \upsilon_n(1, \ldots, 1, v), \qquad \hat{\gamma}_n^{(k)}(u) = \hat{\upsilon}_n^{(k)}(1, \ldots, 1, v),$$

we may conclude that as $n \to \infty$,

$$\left(\alpha_n, \gamma_n, \hat{\alpha}_n^{(1)}, \hat{\gamma}_n^{(1)}, \ldots, \hat{\alpha}_n^{(N)}, \hat{\gamma}_n^{(N)}\right) \rightsquigarrow \left(\alpha, \gamma, \tilde{\alpha}^{(1)}, \tilde{\gamma}^{(1)}, \ldots, \tilde{\alpha}^{(N)}, \tilde{\gamma}^{(N)}\right)$$

in $\mathcal{D}([0,1]^d)^{\otimes 2(N+1)}$, where $\left(\tilde{\alpha}^{(1)}, \tilde{\gamma}^{(1)}\right), \ldots, \left(\tilde{\alpha}^{(N)}, \tilde{\gamma}^{(N)}\right)$ are independent copies of (α, γ), where for any $u, v \in [0,1]^d$, $\alpha(u) = \upsilon(u, 1, \ldots, 1)$ is a C-Brownian bridge and $\gamma(v) = \upsilon(1, \ldots, 1, v)$ is a D-Brownian bridge.

Next, since the conditions of Proposition A.2 of the previous section are met, we obtain that for any $l \in \{1, \ldots, d\}$, $\widehat{\partial_{u_l} C_{n,h}}$ and $\widehat{\partial_{u_l} D_{n,h}}$ converge uniformly in probability to $\partial_{u_l} C$ and $\partial_{u_l} D$.

Hence, defining $\mathcal{E}_n = \mathbb{C}_n - \mathbb{D}_n$ and $\hat{\mathcal{E}}_n^{(k)} = \hat{\mathbb{C}}_{n,h}^{(k)} - \hat{\mathbb{D}}_{n,h}^{(k)}$, it follows that

$$\left(\mathcal{E}_n, \hat{\mathcal{E}}_n^{(1)}, \ldots, \hat{\mathcal{E}}_n^{(N)}\right) \rightsquigarrow (\mathcal{E}, \tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)}) \text{ in } \mathcal{D}\left([0,1]^d\right)^{\otimes(N+1)},$$

where $\tilde{\mathcal{E}}^{(1)}, \ldots, \tilde{\mathcal{E}}^{(N)}$ are independent copies of $\mathcal{E}$. Since the mapping $g \mapsto \int_{[0,1]^d} g^2(u)\mathrm{d}u$ is continuous, whenever g is continuous on $[0,1]^d$, it follows that

$$\left(S_n^{(0)}, \hat{S}_n^{(1)}, \ldots, \hat{S}_n^{(N)}\right) \rightsquigarrow \left(S, \tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}\right) \text{ in } [0,\infty)^{\otimes(N+1)},$$

where $\tilde{S}^{(1)}, \ldots, \tilde{S}^{(N)}$ are independent copies of $S = \int_{[0,1]^d} \mathcal{E}^2(u)\mathrm{d}u$. An approximate p-value for S_n is then given by $\frac{1}{N}\sum_{k=1}^{N} \mathbb{I}\left(\hat{S}_n^{(k)} > S_n\right)$. $\square$

References

[1] P.J. Bickel, D.A. Freedman, Some asymptotic theory for the bootstrap, Ann. Statist. 9 (6) (1981) 1196–1217.
[2] P.J. Bickel, F. Götze, W.R. van Zwet, Resampling fewer than n observations: Gains, losses, and remedies for losses, in: Empirical Bayes, sequential analysis and related topics in statistics and probability (New Brunswick, NJ, 1995), Statist. Sinica 7 (1) (1997) 1–31.
[3] J. Bretagnolle, Lois limites du bootstrap de certaines fonctionnelles, Ann. Inst. H. Poincaré Sect. B (N.S.) 19 (3) (1983) 281–296.
[4] P. Deheuvels, La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendance, Acad. Royal Bel., Bull. Class. Sci., 5e Série 65 (1979) 274–292.
[5] P. Deheuvels, Non parametric tests of independence, in: Nonparametric Asymptotic Statistics (Proceedings of the Conference held in Rouen in 1979), in: Lecture Notes in Mathematics, vol. 821, Springer, New York, 1980, pp. 95–107.
[6] P. Deheuvels, An asymptotic decomposition for multivariate distribution-free tests of independence, J. Multivariate Anal. 11 (1981) 102–113.
[7] P. Deheuvels, A Kolmogorov–Smirnov type test for independence and multivariate samples, Rev. Roum. Math. Pures Appl. 26 (1981) 213–226.
[8] P. Deheuvels, A non parametric test for independence, Publ. Inst. Statist. Univ. Paris 26 (1981) 29–50.
[9] D. Dupuis, N. Papageorgiou, B. Rémillard, E. Jacquier, Empirical study of dependence of credit default data and equity prices, Technical Report G-2006-28, GERAD, 2006.

[10] J.-D. Fermanian, Goodness-of-fit tests for copulas, J. Multivariate Anal. 95 (1) (2005) 119–152.
[11] J.-D. Fermanian, D. Radulović, M.J. Wegkamp, Weak convergence of empirical copula processes, Bernoulli 10 (2004) 847–860.
[12] J.-D. Fermanian, O. Scaillet, Nonparametric estimation of copulas for time series, J. Risk 5 (2003) 25–54.
[13] P. Gänßler, W. Stute, Seminar on Empirical Processes, in: DMV Seminar, vol. 9, Birkhäuser Verlag, Basel, 1987.
[14] C. Genest, K. Ghoudi, L.-P. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika 82 (1995) 543–552.
[15] C. Genest, J.-F. Quessy, B. Rémillard, Goodness-of-fit procedures for copula models based on the integral probability transformation, Scand. J. Statist. 33 (2006) 337–366.
[16] C. Genest, B. Rémillard, Validity of the parametric bootsrap for goodness-of-fit testing in semiparametric models, Ann. Inst. H. Poincaré Sect. B (2008) (in press).
[17] K. Ghoudi, B. Rémillard, Empirical processes based on pseudo-observations. II. The multivariate case, in: Asymptotic Methods in Stochastics, in: Fields Inst. Commun., vol. 44, Amer. Math. Soc, Providence, RI, 2004, pp. 381–406.
[18] K. Grazier, B. G'Sell, Group medical insurance large claims database and collection, in: SOA Monograph M-HB97-1, Society of Actuaries, 1997.
[19] H. Joe, Asymptotic efficiency of the two-stage estimation method for copula-based models, J. Multivariate Anal. 94 (2005) 401–419.
[20] B. Rémillard, The multiplier central limit theorem for pseudo-observations, Technical Report G-2006-30, GERAD, 2006.
[21] B. Rémillard, O. Scaillet, Testing for equality between two empirical copulas, Technical Report G-2006-31, GERAD, 2006.
[22] O. Scaillet, A Kolmogorov–Smirnov type test for positive quadrant dependence, Canad. J. Statist. 33 (3) (2005) 415–427.
[23] O. Scaillet, Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters, J. Multivariate Anal. 98 (2007) 533–543.
[24] J.H. Shih, T.A. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics 51 (1995) 1384–1399.
[25] G.R. Shorack, J.A. Wellner, Empirical processes with applications to statistics, in: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986.
[26] E. Suh, E. Diener, S. Oishi, H. Triandis, The shifting basis of life satisfaction judgments across cultures: Emotions versus norms, Journal of Personality and Social Psychology 74 (2) (1998) 482–493.
[27] H. Tsukahara, Semiparametric estimation in copula models, Canad. J. Statist. 33 (3) (2005) 357–375.
[28] A.W. van der Vaart, J.A. Wellner, Weak convergence and empirical processes, in: Springer Series in Statistics, Springer-Verlag, New York, 1996.
[29] G. van Gurp, G. Meterissian, L. Haiek, J. McCusker, F. Bellavance, St John's wort or sertraline, Canad. Family Phys. 48 (2002) 905–912.
[30] R. Wermers, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, J. Finance 55 (4) (2000) 1655–1695.